



Variant op functiepuntanalyse betrouwbare methodiek

Wat kost dat nou, zo'n datawarehouse?

Harm Schipper

De kosten van datawarehouses blijken moeilijk in te schatten. Menig datawarehouse-project blijkt achteraf aanzienlijk duurder uit te vallen dan in eerste instantie werd begroot. De oorzaak van dit probleem ligt vaak in de onervarenheid van organisaties met het bepalen van de omvang in werkdagen van een datawarehouse-project.

Schattingen worden met de 'natte vinger' gemaakt en veelal gebaseerd op ervaringen met het bouwen van operationele systemen. Maar het bouwen van een datawarehouse kan geenszins vergeleken worden met het bouwen van een operationeel systeem. Een operationeel systeem is bedoeld voor het efficiënt uitvoeren van dagelijkse operationele taken en is daarom geoptimaliseerd voor transactieverwerking. Het onderliggende datamodel is relationeel ingericht. Een datawarehouse daarentegen is bedoeld voor rapportage en analyse van grote hoeveelheden informatie. Een datawarehouse is daarom geoptimaliseerd voor het binnenhalen van grote hoeveelheden data uit bronsystemen, het opbouwen van historie en het leveren van informatie. Het onderliggende datamodel is dimensioneel ingericht.

De gegevensverzamelingen kunnen op drie manieren worden gebruikt

Het gevolg van foutieve inschattingen betreffende datawarehouse-projecten is meestal een ongewenste budgetoverschrijding en het afraffelen van de bouw tegen het einde van het project. Vaak wordt in het eindstadium besloten specifieke functionaliteiten toch maar niet te realiseren. De eindgebruikers zijn hier uiteraard de dupe van. Het uiteindelijke gevolg is namelijk dat een organisatie wordt opgescheept met een (te) duur datawarehouse, dat niet optimaal gebruikt kan worden. Dit heeft in menige organisatie geleid tot scepsis jegens het datawarehouse. Alleen al het noemen van het woord datawarehouse kan tot geërgerde reacties lijden. Hoe komt men tot een betere inschatting van de omvang van een datawarehouse-bouwproject? Bij reguliere ontwikkeltrajecten

worden inschattingen meestal gemaakt door het werk op te delen in een aantal componenten die wat betreft functie overeenkomen. Voor iedere component is dan een standaardmethode beschikbaar om de benodigde inspanning voor de te realiseren component te bepalen. De functiepuntanalyse is een voorbeeld van een dergelijke methode; met een variant op de functiepuntanalyse kunnen ook de omvang en kosten voor het bouwen van een datawarehouse worden berekenend.

Functiepuntanalyse

Functiepuntanalyse is een methode om tot een objectieve en technologie-onafhankelijke inschatting te komen van de grootte van een informatiesysteem¹. De Nederlandse Software Metrieken Gebruikers Associatie (NESMA, www.nesma.nl) is de organisatie die het gebruik van functiepuntanalyse in Nederland bevordert. Functiepuntanalyse wordt gedaan vanuit het gezichtspunt van de gebruikers, dat wil zeggen vanuit een logisch, niet-technisch oogpunt. Met gebruikers worden in dit verband niet alleen de eindgebruikers bedoeld, maar bijvoorbeeld ook de beheerder van een informatiesysteem of een andere applicatie die data opvraagt uit het systeem.

Voor het uitvoeren van een functiepuntanalyse wordt het te bouwen informatiesysteem geanalyseerd op gebruikersfuncties. Het overzicht van gebruikersfuncties wordt opgedeeld in gebruikersprocessen en gebruikersgegevensverzamelingen. Vanuit het gezichtspunt van de gebruiker is een gebruikersgegevensverzameling een bij elkaar horende groep gegevens. Dit kunnen systeemeigen en systeemvreemde gegevens zijn. Systeemvreemde gegevensverzamelingen worden door een ander informatiesysteem onderhouden.

De gegevensverzamelingen kunnen op drie manieren worden gebruikt; voor het invoerproces, het uitvoerproces en het opvragingsproces. Tot het realiseren van een gegevensverzameling of

proces hoort niet alleen het bouwen maar ook zaken als testen, managen en documenteren. Functiepuntanalyse leidt tot een overzicht van deze gebruikersfuncties, de complexiteit van de functies en de zwaarte van de systeemkenmerken. Het resultaat wordt uitgedrukt in functiepunten.

Het optellen van alle functiepunten resulteert in het aantal bruto functiepunten (BFP). Op het BFP wordt vervolgens een correctiefactor toegepast voor zaken als installatiegemak, bedieningsgemak, transactievolume etcetera. Dit is een standaard correctiefactor, die is bepaald door het NESMA, en resulteert in het aantal netto functiepunten (NFP). Door toepassing van de productiviteitsfactor (uitgedrukt in aantal uren per functiepunt) wordt tenslotte het aantal uren verkregen dat nodig is voor het bouwen van het informatiesysteem. Deze productiviteitsfactor is afhankelijk van de complexiteit van de tools die worden gebruikt en de ervarenheid van de werknemers die aan het informatiesysteem werken. Door tijdens de uitvoering van het project de daadwerkelijk gemaakte uren bij te houden, kunnen precieze productiviteitsfactoren worden verkregen. De functiepuntanalyse is in de praktijk een betrouwbare methode gebleken voor het schatten van de omvang van een informatiesysteem.

Model-datawarehouse

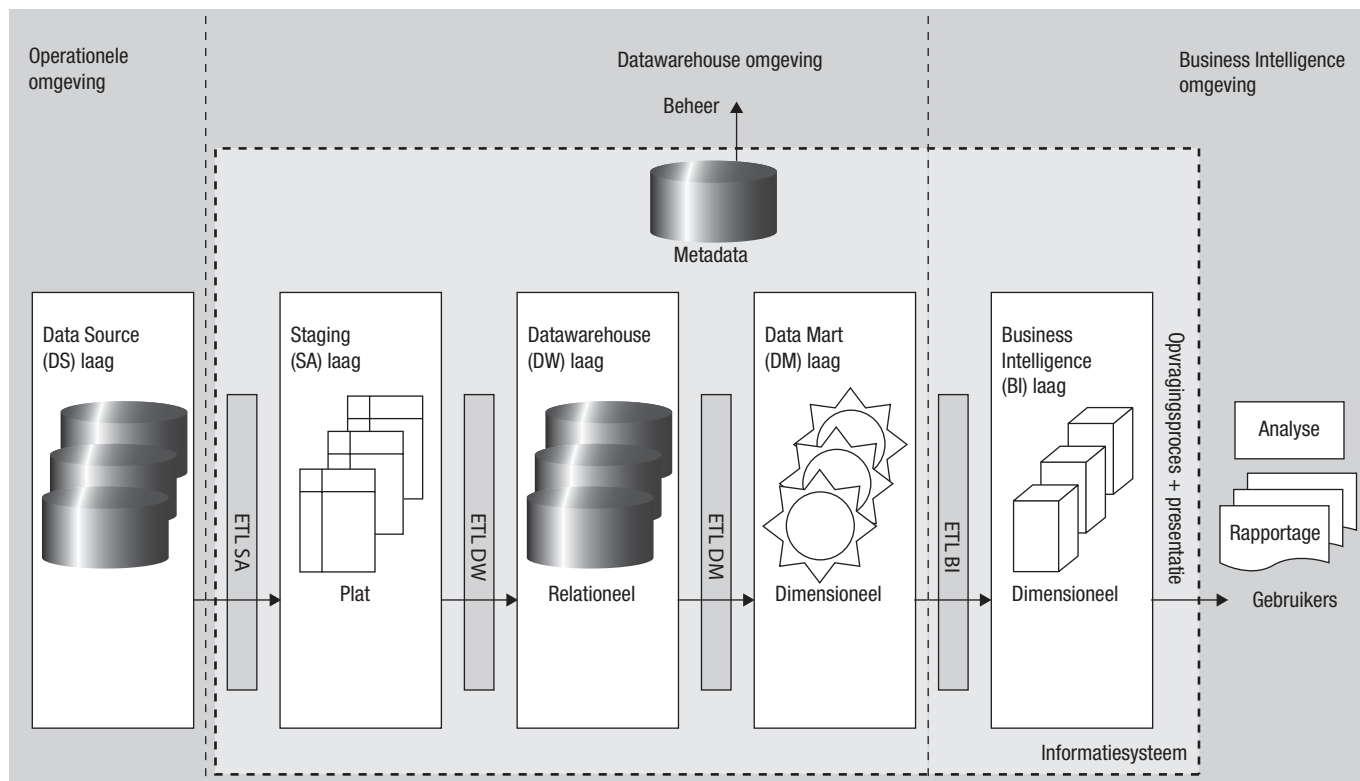
De bouw van datawarehouse-informatiesystemen verschilt in belangrijke mate met de bouw van operationele informatiesystemen. Datawarehouse-specifieke zaken als sterschema's, kubussen, ETL-processen en metadata zijn dan ook niet zonder

meer in te schatten met de standaard functiepuntanalyse-methodiek. Op basis van de ervaring opgebouwd met het bouwen van datawarehouses en het uitvoeren van de gebruikelijke functiepuntanalyses is een datawarehouse-variant op de functiepuntanalyse (FPA-dwh) ontwikkeld. Deze methodiek is onderdeel een geoptimaliseerde software engineering-methodiek, genaamd BI factory, voor Business Intelligence en is gericht op het versneld leveren van oplossingen van betrouwbare kwaliteit.

Aan de hand van een model-datawarehouse wordt gedemonstreerd hoe de kostprijs voor het bouwen van een datawarehouse wordt berekend. Het model is weergegeven in afbeelding 1. In de praktijk zullen uiteraard allerhande varianten op dit model te vinden zijn. Dit is geen probleem; de kostprijsberekening is eenvoudig aan te passen aan iedere variant. In de afbeelding zijn tevens de grenzen van het informatiesysteem (gestippelde vierkant) en de gebruikers aangegeven. De gebruikers van het datawarehouse zijn ook in dit geval niet alleen de eindgebruikers, die rapporten ontvangen en data uit de datawarehouse analyseren, maar ook datawarehouse-beheerders en bronsystemen die data aanleveren.

In het model zijn de drie omgevingen aangegeven die altijd in een datawarehouse aanwezig zijn:

1. Operationele omgeving: de omgeving met de operationele systemen die als bronnen van informatie voor het datawarehouse fungeren;
2. Datawarehouse-omgeving: de omgeving waarin de data worden opgeslagen en historie wordt opgebouwd;
3. Business Intelligence-omgeving: in de BI-omgeving wordt de



Afbeelding 1: Een voorbeeld van een datawarehouse-ontwerp.

ID Component	Tool	Aantal BFP	Correctie-factor	DWH correctie-factor	Aantal NFP	Productiviteitsfactor (uur/NF)	Aantal uren
1 DS laag	Diverse	56	1.1	0.98	60	3	181
2 ETL SA	ETL Tool	32	1.1	0.98	34	10	345
3 SA laag	DB Tool	67	1.1	0.98	72	3	217
4 ETL DW	ETL Tool	123	1.1	0.98	133	15	1989
5 DW laag	DB Tool	224	1.1	0.98	241	3	724
6 ETL DM	ETL Tool	23	1.1	0.98	25	15	372
7 DM laag	DB Tool	47	1.1	0.98	51	3	152
8 ETL BI	ETL Tool	55	1.1	0.98	59	12	711
9 BI laag	BI Tool	105	1.1	0.98	113	6	679
10 Analyse & rapportage	BI Tool	236	1.1	0.98	254	8	2035
11 Metadata	DB Tool	32	1.1	0.98	34	3	103
Totaal		1000			1078		7509
Stap		1			2		3

BFP = Bruto Functiepunten; BI = Business Intelligence; DB = Database; DM = Datamart; DWH = Datawarehouse; ETL = Extraction Transformation & Loading; NFP = Netto Functiepunten

Afbeelding 2: Functiepunctanalyse voor het model datawarehouse uit afbeelding 1 (getallen en factoren zijn fictief).

informatie opgeslagen in het datawarehouse beschikbaar gesteld aan de eindgebruikers voor rapportage en analyse.

In de FPA-dwh methodiek wordt de datawarehouse opgedeeld in een aantal componenten. Deze componenten zijn genummerd weergegeven in afbeelding 1. Voor iedere component geldt een specifieke berekeningswijze voor het bepalen van het aantal functiepunten, waarbij gebruik wordt gemaakt van dezelfde gebruikersfuncties als in de standaard functiepunctanalyse.

Het berekenen van de kostprijs van een datawarehouse is een proces van drie stappen:

1. Berekening van inspanning in uren voor het realiseren van de datawarehouse;
2. Uitdrukken totale inspanning in verschillende taken en rollen van het ontwikkel-team;
3. Het berekenen van de uiteindelijke kostprijs van de datawarehouse op basis van de uitkomsten van bovenstaande stappen.

Stap 1: bepalen omvang van het datawarehouse

In stap 1 wordt met behulp van FPA-dwh het totaal aantal uren bepaald voor de vervaardigen van de datawarehouse. Een korte toelichting op de FPA-dwh methode volgt hierna.

Functiepunctanalyse voor de DS-laag.

Er kunnen zich drie situaties voordoen in de DS-laag. Iedere situatie heeft consequenties voor het bepalen van het aantal functiepunten:

1. De bronbestanden worden kant en klaar door de aanleverende partij(en) aan het datawarehouse aangeleverd. In dit geval worden geen functiepunten berekend voor de DS laag.

2. De bronbestanden worden vanuit het datawarehouse aangemaakt. In dit geval wordt ieder bronbestand beschouwd als een systeemvreemde gegevensverzameling. Het aantal recordtypen is het aantal tabellen in de operationele systemen, waaruit de data worden gelezen, en het aantal data-elementtypen is het aantal attributen in het bronbestand. In de functiepunctanalyse wordt met recordtype het aantal entiteitstypen bedoeld en met data-elementtype de meest elementaire vorm van een gegeven zoals die door de gebruiker wordt beschouwd.
3. De bronbestanden worden in de aanleverende systemen aangemaakt. In dit geval wordt ieder aangemaakt bronbestand als een uitvoerproces beschouwd.

Functiepunctanalyse voor de diverse ETL-componenten.

Voor het berekenen van het aantal functiepunten voor een ETL-component wordt voor iedere tabel in de doellaag een invoerproces geteld omdat de ETL in principe steeds een uniek proces is. Met tabel wordt in de SA-laag daadwerkelijk een tabel bedoeld, in de DW-laag een systeemeigen gegevensverzameling, in de DM-laag een sterschema en in de BI-laag een kubus. Voor een ETL-proces is het aantal systeemeigen gegevensverzamelingen het aantal tabellen waaruit de data worden gehaald in de bron-laag. Het aantal data-elementtypen is het aantal attributen dat daadwerkelijk wordt ingelezen plus het aantal attributen dat in de doellaag wordt geladen.

Functiepunctanalyse voor de SA-, DW-, DM- en BI-laag componenten.

Voor de SA- en DW-laag gelden de normale telrichtlijnen van NESMA. Voor de DM- en BI-laag worden sterschema's en kubussen beschouwd als systeemeigen gegevensverzamelingen. Het aantal recordtypen is daarbij gelijk aan het aantal dimensie-

tabellen plus één (voor de feitentabel) en het aantal data-elementtypen is gelijk aan het aantal niveaus van de dimensie plus een voor de waarde.

Functiepuntnalyse voor opvragingsproces (10 in afbeelding 1).

De component van het opvragingsproces kan uit één of meerdere functies bestaan:

1. Aanmaken van standaard rapporten die gebruikers kunnen kiezen en uitprinten;
2. Aanmaken van overzichten waarin gebruikers meetwaarden en dimensies kiezen;
3. Applicaties voor analyseren van data met behulp van allerhande OLAP functies.

Voor de functies 1 en 2 kunnen in principe de NESMA-telrichtlijnen worden ahouden. Het tellen van functie 3 is afhankelijk van het feit of hiervoor een standaard BI-tool is aangeschaft of dat allerhande OLAP-functies zelf worden gebouwd. Maar ook in het aanbod van BI-tools is veel variatie. Zo kan de gebruiker met de ene tool direct na installatie aan de slag maar is voor een andere tool nog veel ontwikkelwerk vereist voordat de data daadwerkelijk kunnen worden geanalyseerd.

Functiepuntnalyse voor de metadata-component.

Voor de metadata-component kunnen de standaard NESMA telrichtlijnen worden gevolgd. Door voor iedere component in het datawarehouse het aantal functiepunten te bepalen en deze bij elkaar op te tellen, wordt het totaal BFP verkregen. Zoals gebruikelijk in de standaard functiepuntnalyse wordt ook voor datawarehouses het BFP gecorrigeerd met de NESMA-correctiefactor. In FPA-dwh wordt echter nog een extra correctiefactor geïntroduceerd: de datawarehouse-correctiefactor. Deze factor bestaat uit specifieke datawarehouse-elementen waarvoor het BFP moet worden gecorrigeerd.

Een voorbeeld is de mate van vervuiling van de brondata. Als de brondata bovenmatig vervuild zijn met dubbele of foutieve invoerwaarden en dergelijke, en een meer dan gemiddelde inspanning vereist is om de data te schonen, kan dat leiden tot een hogere correctiefactor en daardoor een hoger aantal functiepunten. Het resultaat na correctie van de BFP is het totaal NFP. Het NFP is een goede maat voor de omvang van een datawarehouse maar geeft nog niet aan welke inspanning is vereist om het datawarehouse te realiseren in termen van aantal uren. Hiertoe wordt het NFP vermenigvuldigd met de productiviteitsfactor. In afbeelding 2 is een voorbeeld gegeven van een functiepuntnalyse voor het model-datawarehouse uit afbeelding 1. Uiteraard zijn de aantallen in het voorbeeld verzonden.

Stap 2: verdelen van taken en rollen

Op basis van het aantal uren berekend in stap 1 en de gehanteerde methodiek voor uitvoering van het project (de projectaanpak van Capgemini is gebaseerd op DSDM en RUP) kan het aantal uren worden verdeeld over de verschillende projecttaken

en -rollen. In afbeelding 3 is hiervan een voorbeeld gegeven. Hierbij is uitgegaan van de DSDM/RUP projectaanpak en de rollen van projectmanager, ontwerper, architect, test-analist, analist en ontwikkelaar.

Afbeelding 3 gaat uit van het realiseren van het project in eigen beheer en op de eigen locatie. Maar het model kan eenvoudig worden uitgebreid als het project deels of geheel wordt uitbesteed. Dit uitbesteden kan onshore (zelfde land maar op een andere locatie) of zelfs offshore (een lage-lonen land bijvoorbeeld) gebeuren. Hierbij moet vastgesteld worden welk deel van iedere taak onsite, onshore en offshore wordt uitgevoerd en welke rollen daarbij op iedere locatie zijn betrokken. Momenteel is er een duidelijke trend waarneembaar naar het offshore uitvoeren van datawarehouse-projecten. Het gaat hierbij vooral om het uitbesteden van ontwikkelwerk naar landen als India. Bij het (deels) offshore uitvoeren van projecten moet men in het model rekening houden met extra inspanningen die samenhangen met juist het werken volgens het rightshore² model.

Stap 3: wat kost het?

Nu de totale inspanning is uitgerekend en de taken over de diverse rollen zijn verdeeld, kan op basis van (interne) tarieven de uiteindelijke kostprijs van een datawarehouse worden bepaald. Ook in dit geval is het model eenvoudig uit te breiden indien de opdracht deels wordt uitbesteed en/of volgens het rightshore model wordt uitgevoerd. Uiteraard zijn de tarieven op de diverse locaties niet hetzelfde en daarbij moet rekening worden gehouden met tariefopslagen voor zaken als reizen en communicatie.

Conclusie

Datawarehouse-projecten zijn lastige projecten en hebben de neiging budget-limieten te overschrijden. Door een datawarehouse te ontleden in een aantal componenten, per component op objectieve wijze de omvang te bepalen en deze te vermenigvuldigen met een, op ervaringen gebaseerde, productiviteitsfactor, krijgt men een betrouwbare schatting van de kosten van het project. Heeft u het plan opgevat om een datawarehouse te gaan bouwen of een datamart aan een bestaande datawarehouse toe te voegen, dan doet u er verstandig aan een realistische kosten-calcuatie te maken op basis van de datawarehouse-variant van de functiepuntnalyse-methode. U weet dan precies waaraan u begint en u hoeft niet achteraf te concluderen dat het project aanmerkelijk duurder is geworden dan van te voren gedacht.

Noten

1. *Definities en telrichtlijnen voor de toepassing van functiepuntnalyse. Versie 2.2, 2004. Nederlandse Software Metrieken Gebruikers Associatie (NESMA).*
2. *Rightshore is een door Capgemini gepatenteerde term en betekent voor een software-project het geven van de juiste service op de juiste plaats en voor de juiste prijs.*

Harm Schipper

Dr. Harm Schipper (harm.schipper@capgemini.com) is werkzaam als consultant bij Capgemini.